

Рекомендательные системы в публичных библиотеках

Александр Сергеевич Карауш

Муниципальная информационная библиотечная система г.Томска

В последнее время отмечается существенное ослабление интереса у населения к досуговому чтению и проведению свободного времени в виде чтения современной художественной или научно-популярной литературы. Многие читатели объективно полагают, что современные возможности библиотек не позволят им найти документы или данные, с помощью которых можно с пользой провести свое личное время. Намного проще использовать «подготовленный» для восприятия видео-контент или звуковой контент. Для таких решений потенциальных читателей имеются следующие предпосылки:

- взрывной рост объемов доступных обществу данных (увеличение числа книг, фильмов, новостей, рекламных сообщений и пр.);
- увеличение объема онлайн-данных;
- реальный объем информации, окружающей человека, значительно выше того, что он может реально пропустить через себя, чтобы обнаружить необходимую и достаточную, а также понравившуюся ему.

Решения о фильтрации поступающих данных тесно связаны с технологиями:

1. Нужна технология, которая помогла бы читателю найти то, в чём он нуждается, и избежать того, на что не желает тратить своё время и внимание.
2. Для решения проблемы "информационного вала" часто используется метод фильтрации контента, который необходимо совершенствовать.
3. Для пользовательского потребления система должна рекомендовать объекты и данные на основе корреляции содержания с пользовательскими предпочтениями. Например, система может попытаться установить корреляцию между наличием ключевых слов в статье и пользовательскими предпочтениями.

Сегодня библиотеки разных ведомств и уровней пытаются разобраться с показателями собственной деятельности и определить границы применимости этих показателей для новых сред распространения данных. Показатели (или индикаторы) работы библиотеки должны быть, прежде всего: объективные; легко вычисляемые; «конвертируемые» в «небиблиотечные» индикаторы и показатели (финансовые, кадровые, ресурсные и пр.); относительно «стабильные» во времени.

Для публичных и массовых библиотек особенно важно внимание населения к досуговому и развивающему чтению, поскольку это позволит библиотекам формировать и развивать пользовательские группы (целевые аудитории для потребления услуг предоставления информации).

Автор придерживается мнения, что основным критерием работы современной публичной библиотеки является **качественное личностное время** пользователя - время, которое человек субъективно оценивает как хорошо и полезно проведённое.

Основной показатель (индикатор) работы публичной библиотеки - количество качественного личностного времени, которое читатели приобрели, используя ресурсы конкретной библиотеки (книги, комфорт, тепло, цифровые данные, аудио- и видеодокументы и т.д.), присутствуя там непосредственно или дистанционно (удаленно).

В этой связи стоит рассмотреть наработки в библиотечном деле по направлению развития сотрудников до уровня **«сетевого библиотекаря»** или **«библиотекаря 2.0»**. В настоящее время для увеличения качественного личностного времени читателя библиотеками активно используются следующие технологии и сервисы:

- Интернет-справка и пр. (вопрос-ответ, форумы, блоги и пр.);
- создание профилей пользователей (скрытое и явное);
- фильтрация данных поиска (Интернет, пиринговые сети и пр.):
 - автоматические системы рекомендаций;
 - индивидуальные рейтинги (в т.ч. без возможности общения);

- развитие картотек пользователей библиотеки в АБИС до уровня CRM (система управления пользователями).

Коллаборативная фильтрация [1, 4] может и должна использоваться в АБИС, как основная для фильтрации передаваемых данных между АБИС и пользователем.

Коллаборативная фильтрация, совместная фильтрация (collaborative filtration) - метод, дающий автоматические прогнозы (фильтрацию) относительно интересов пользователя по предварительно собранным данным о вкусах множества пользователей, сотрудничающих между собой. Где определяется, какие из пользователей имеют тот же вкус на основании стандартных формул вычисления статистических корреляций, при этом основное допущение: те, кто соглашался в прошлом, склонны согласиться и в будущем.

Коллаборативная фильтрация отличается от рейтинга, дающего усреднённую оценку для каждого объекта интереса, базирующуюся на количестве поданных за него голосов.

Сегодня уже можно описать круг задач и областей применения для автоматизированных библиотечных систем, основанных на использовании рекомендательных сервисов, в частности, коллаборативной фильтрации [2, 5]:

- **Аннотация в контексте.** Фильтрация сообщений на структурированных форумах и блогах с целью принятия решения, какие из них стоит читать.
- **Найти некоторые хорошие объекты.** Пользователю предлагаются конкретные объекты, представленные в виде ранжированного перечня рекомендуемых объектов вместе с прогнозом, насколько сильно они ему понравятся.
- **Найти ВСЕ хорошие объекты.** Позволяет выделить из массива объектов не несколько хороших, а ВСЕ.
- **Последовательность рекомендаций** для пользователя из нескольких объектов. При чтении (изучении) каждого из них поочередно пользователь получит качественное личностное время, переходя от одного объекта (например менее сложного в понимании) к другому (более сложному), и развиваясь как личность.
- **Только просматривание.** Для кого-то это развлечение, для кого-то - процесс познания группы единомышленников.
- **Поиск пользователем надёжной системы** коллаборативной фильтрации. Пользователи "забавляются" с системой какое-то время, чтобы посмотреть, удовлетворяют ли рекомендуемые объекты их вкусам с учетом изменения ими поставленных ранее оценок для объектов.

Типовой пример алгоритма рекомендательной системы на основе коллаборативной фильтрации [1, 3]:

1. Каждый пользователь в некоторой группе оценивает определенное количество объектов потребления, например, книг.
2. Каждой книге ставится оценка: «Очень понравилось», «Понравилось», «Так себе», «Скучновато» и «Плохо».
3. На основании данных о предпочтениях для каждого пользователя формируется круг единомышленников, вкус которых по отношению к оцененным им книгам совпадает с его личными пристрастиями.
4. Выбираются произведения, прочитанные и высоко оцененные единомышленниками, но не прочитанные данным человеком – эти произведения и составляют выданные ему рекомендации.

В результате лучше предсказывается оценка человеком неизвестных ему объектов, чем следование рейтингам и общесоциологическим данным: пол, возраст, образование, социальный статус и т.д.

При использовании вышеприведенного примера алгоритма можно уже сегодня получить новые сервисы для пользователей библиотек, проведя доработки в АБИС:

1. рекомендательные списки к чтению для **всех видов носителей и документов**;
2. составление **индивидуальных рейтингов и списков** (последовательность рекомендуемых объектов);

3. навигация на основе предоставленных рекомендаций;
4. **индивидуальное ранжирование результатов поиска** в АБИС.

Кроме очевидных достоинств стоит упомянуть и про имеющиеся недостатки рекомендательных систем, построенных с использованием модели коллаборативной фильтрации:

1. **Проблема накопления «личного отрицательного опыта»** связана с чтением не всегда «приятных» для пользователя объектов и анализом отрицательной информации. Система коллаборативной фильтрации рекомендует только понравившиеся объекты для чтения.
2. **Рекомендуются документы уже кем-то прочитанные и оцененные.** Пользователю нет возможности рекомендовать книгу, на которую нет отзыва в системе.

Новые задачи определяют проблемы и недоработки АБИС, которые необходимо решить для использования новых возможностей обслуживания пользователей:

1. Развитие АБИС по законам и направлениям развития CRM.
2. Использование весовых коэффициентов для каждого возвращенного объекта при поиске.
3. Обмен данными с другими современными CRM: единый формат обмена; формат профиля пользователя; особенности работы с «личной информацией».
4. Автоматически просчитываемые индивидуальные рейтинги для чтения.
5. Автоматические анализаторы спроса на литературу для определения объемов предполагаемого комплектования библиотеки.
6. Технологии передачи рекомендаций на объекты из других систем: электронные библиотеки; Интернет-магазины, образовательные стандарты, списки победителей конкурсов и пр.
7. Использование сетевой модели баз данных для хранения и пересчета множества взаимных коэффициентов для связей пользователей между собой, а также для объединения их в группы «по интересам».

Выводы

Развитие современных технологий получения данных заставляет библиотеки искать новые подходы для определения значимости в глазах пользователя. Попытки следовать только лишь в одном направлении – мемориальном, в настоящее время показали свою несостоятельность по причине уменьшения стоимости носителей информации и технологий хранения. Почти забытая «библиотечная технология» - рекомендовать непросвещенным пользователям новые книги (ссылки, файлы и пр.), с использованием новых информационных технологий позволяет создать новые возможности для пользователей библиотек и потенциальных читателей.

Литература

1. Долгин А. Экономика символического обмена. – М.: Инфра-М, 2006. – 632 с.
2. Herlocker J.L., Konstan J.A., Terveen L.G., Riedl J.T. Evaluating Collaborative Filtering Recommender Systems // ACM Transactions on Information Systems, Vol. 22, № 1, 2004. P. 5-53.
3. Herlocker J.L., Konstan J.A. Content-Independent Task-Focused Recommendation // IEEE Internet Computing, Vol. 5, № 6, November/December, 2001. P. 40-47.
4. O'Mahony M., Hurley N., Kushmerick N., Silvestre G. Collaborative recommendation: A robustness analysis // ACM Transactions on Internet Technology, Vol. 4, Issue 4, November 2004. P. 344-377.
5. Adomavicius G., Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions // IEEE Transactions on Knowledge and Data Engineering, Vol. 17, № 6, June, 2005